Homework Assignment # 3

Due: Thursday, April 4, 2019, 11:59 p.m.

Total Points: 50

This assignment is designed to help you better understand concepts that were presented during class. You may work in small groups (2 or 3 individuals) on the homework assignments, but it is expected that you submit your own work. Therefore, **each student is responsible for submitting their own homework**, whether they work in groups or individually. The submitted homework must include the name of other individuals you worked with (if applicable), as well as all resources that were used to solve the problem (e.g. web sites, books, research papers, other people, etc.). Academic honesty is taken seriously; for detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

Your assignment must be submitted as a single pdf document on Canvas. The questions and your answers must be typed; for example, in Latex, Microsoft Word, Lyx, etc. Images may be scanned and inserted into the document if it is too complicated to draw them properly. Be sure to show all work.

All code (if applicable) should be turned in when you submit your assignment. Use Matlab, Python, R, Java or C. Be sure to include an complete explanation on how to run/execute your code.

All assignments must be submitted on time to receive credit. No late work will be accepted, unless you have a prior arrangement with the instructor.

Question 1. [10 POINTS]

The standard linear regression model is: $y = Xw + e$, where X is an $n \times d$ matrix of predictor variables, y is an $n-$dimensional vector of response variables, and $e \sim \mathcal{N}(0, \sigma^2 I)$ is an $n-$dimensional vector of model errors.

(a) What is the PDF of y in terms of X, w, σ^2?

(b) Let the PDF from part (a) be denoted as $f(y|w)$. Suppose also in this case that $w \sim \mathcal{N}(0, \rho^2 I)$. Write an expression for the joint PDF of w and y (e.g. $f(w, y)$).

(c) Show that

$$\hat{w}_{MAP} = \arg\min_{w} \frac{||w||^2}{\rho^2} + \frac{||y - Xw||^2}{\sigma^2} \tag{1}$$

Question 2. [10 POINTS]

Consider the standard linear regression model: $y = Xw + e$, where X is an $n \times d$ matrix of predictor variables and y is an $n-$dimensional vector of response variables. Let the error terms be i.i.d. with the following distribution $e_1, e_2, \ldots, e_n \sim \mathcal{N}(\mu, \sigma^2)$. Derive an unbiased estimate for μ in this model.

Question 3. [10 POINTS]

Two data sets are posted on Canvas. The first has an $n \times d = 1000 \times 50$ data matrix (X) "pred1.dat" with a 1000×1 response vector (y) in "resp1.dat." The second has a 1000×500 data matrix "pred2.dat" with a response vector in "resp2.dat." These data sets were generated according to the standard linear regression model, as described in problem 1, but where the weights, w are NOT random variables.

(a) For each data set, use the first half of the data (observations $i = 1, \ldots, n/2$, all d predictors) to estimate values for w, $\hat{w}$.

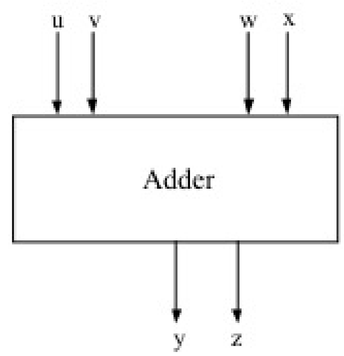
(b) For each data set, use your estimate of w on the 2nd half of the data set $(n/2+1, ..., n)$, to get your estimated response variables, $\hat{y}$. Compute and report your total squared error: $SSE = \sum_{i=n/2+1}^{n} (\hat{y}_i - y_i)^2$

Be sure to include the results and your code in the solutions report.

Question 4. [10 POINTS]

For this problem, change the definition of an M-P neuron so that both its inputs and outputs are binary (0 or 1). View uv and wx each as two-bit binary (00, 01, 10, or 11) numbers, and yz as the 2 low-order bits of the numerical addition of uv and wx (see the figure below).

(a) Give weights and biases for an M-P network which generates z.
(b) Give weights and biases for an M-P network which generates y.



Question 5. [10 POINTS]

For the following training samples:

$\mathbf{x}_1 = (0,0)^T \in C_1$
$\mathbf{x}_2 = (0,1)^T \in C_1$
$\mathbf{x}_3 = (1,0)^T \in C_2$
$\mathbf{x}_4 = (1,1)^T \in C_2$

(a) Plot them in input space.
(b) Apply the perceptron learning rule to the above samples one-at-a-time to obtain weights that separate the training samples. Set η to 0.5. Work in the space with the bias as another input element. Use $\mathbf{w}(0) = (0,0,0)^T$. Show values for $\mathbf{w}$ after it is updated for each training sample.
(c) Write the expression for the resulting decision boundary from (b).
(d) XOR. For $\mathbf{x}_2, \mathbf{x}_3 \in C_1$ and $\mathbf{x}_1, \mathbf{x}_4 \in C_2$, describe your observation when you apply the perceptron learning rule following the same procedure as in (a)-(c).